

Lexical Phonology and the Rebirth of the Phoneme*

Paul Kroeger
9-May-90

In the past 15 years or so, a number of traditional concepts have been re-introduced into current phonological theory. One could point to the development of metrical phonology based on the traditional notions foot and meter, work in prosodic phonology based on a hierarchy of phonological units, the autosegmental development of Firthian “prosodies”, etc.

This paper examines another such example, the re-emergence (in the theory of Lexical Phonology) of a level of representation corresponding in many ways to the classical phonemic level of American Descriptivism. The analogy between the descriptivists’ phonemic level and the output of the lexical component in Lexical Phonology (LP) has been frequently noted. The goal of the present paper will be to clarify the differences as well as the similarities between the two, and to give some historical perspective on the issues involved.

As Mohanan (1986) points out, the core intuition behind the definition of the phoneme was that speakers of a language react to some phonetic distinctions while ignoring others:

Broadly speaking, the classical phonemic level of representation arose out of the speaker’s intuitions about what he was saying or hearing, or what was significant in it. This level was meant to capture the speaker’s intuitions about which sounds were the same or different: thus, speakers of English judge [t] and [t^h] to be the same, while they judge [t] and [s] to be distinct.

*Many of the ideas in this paper are drawn from classroom lectures by Paul Kiparsky and Don Burquest. My thanks to D. Burquest, K.P. Mohanan and K.L. Pike for comments on an earlier draft of this paper.

... Lexical Phonology tries to regain what was intuitively true about the classical phonemic representation. (pp. 6-7)

The crucial difference between the descriptivist phonemic level and the level of “Lexical Representation” in LP is the requirement in the former of a biunique mapping between phonemic and phonetic representations. This difference in turn follows from basic differences in the theoretical assumptions of the two frameworks. The descriptivists sought to define an inventory of units, and to describe the distribution of these units. If two units (e.g. two phonemes) were shown to be significantly different in one environment, they had to be considered distinct in all environments.

Lexical Phonology, on the other hand, seeks to determine a system of rules which will determine the surface forms of a language in a maximally efficient and elegant way. Rather than distinguishing between two kinds of units (phonemes vs. morphophonemes), LP distinguishes between two modes of rule application: lexical (word-level) vs. post-lexical (phrase-level or syntactic). The classical phonemic level corresponds roughly to the output of the word-level phonology and the input to the phrase-level phonology.

1. Overview of Lexical Phonology

Some phonological processes apply strictly within words, while others may apply either within words or across word boundaries. The basic claim of LP is that there are certain characteristic properties that distinguish word-bounded rules from non-word-bounded rules, and that all word-bounded rules must apply before the non-word-bounded rules.

The lexicon is viewed as the component of the grammar in which words are formed. Both morphological and phonological processes take place within this component. Within the LP model, these morphological and phonological processes are “interleaved”: some

phonological rules (lexical rules) may apply after each successive morphological operation. This is a radical departure from previous (and many current) theories, which assume that all morphological processes must precede all phonological rules.

The output of the lexicon is a word. Rules of syntax combine words into phrases, sentences, etc. after which another set of phonological rules (post-lexical rules) may apply. In many cases the same rule may apply both lexically and post-lexically, but the application of the rule is subject to different constraints in the two modules.

Kaisse and Shaw (1985) summarize the differences between lexical and post-lexical applications of phonological rules as follows:

- (1) Lexical rules never apply across word boundaries, whereas post-lexical rules may apply both within words and across word boundaries.
- (2) Post-lexical rules are exceptionless -- they apply wherever their structural description (i.e. conditioning environment) is satisfied. Lexical rules, however, often have marked exceptions.
- (3) Lexical rules are structure preserving, i.e. they do not create segments not present in underlying forms in that language. Post-lexical rules may create “novel” segment types. Another way of stating this is that lexical rules change one “phoneme” into another -- they cannot modify non-distinctive features. Post-lexical rules, on the other hand, may modify both distinctive and non-distinctive features.
- (4) Lexical rules are categorical, but post-lexical rules may produce gradient (non-binary) outputs.
- (5) Lexical rules apply only in derived environments (e.g. across morpheme boundaries), while post-lexical rules may apply in underived environments (e.g. morpheme-internally).

(6) Post-lexical rules can never apply cyclically, whereas most lexical rules are cyclic. Post-lexical rules can not be sensitive to morphological structure.

Strictly speaking, rules themselves cannot in general be classified as being "lexical" or "post-lexical". As noted above, a single rule may apply both lexically and post-lexically. The term "lexical rule" is thus a short-hand way of referring to the application of a given rule in the lexical component of the phonology. Counter-examples to several of the generalizations listed above have been proposed in the literature; see, for example, Mohanan and Mohanan (1984). But even those that have been challenged hold at least as strong tendencies.

Within the lexicon, morphological processes often seem to cluster together both in terms of relative position of affixation and in the phonological rules which apply to the output of each morphological process. Such clusters define "strata" or levels of morphology, which are assumed to be linearly ordered with respect to each other. Lexical phonological rules may be restricted to apply only in specific strata.

For example, the agentive/instrumental suffix -er in English has different phonological effects than the homophonous comparative suffix. The /g/ in stems ending in /-ng/ is silent before the agentive suffix, just as it is in word-final position: singer, hanger, stinger, bringer, etc. The same pattern holds in compound words: long-eared, hangout, singalong, etc. However, before the comparative -er the /g/ is pronounced, as in morpheme-internal /-ng-/ clusters: longer, stronger, etc.; compare hunger, finger, anger, etc.

Sapir (1925), to whom this contrast was pointed out by Bloomfield, said: "[Agentive] -er might almost be construed as a 'word' which occurs only as the second element of a compound ...". In Lexical Phonology, this same intuition is captured by assigning the comparative and agentive suffixes to different strata. The comparative suffix would be added in the first stratum,

the agentive suffix in the second stratum, the same stratum where Kiparsky (1982) suggests that compounds are formed. The rule of final /g/-deletion in /ng/ clusters could apply in stratum 2 and post-lexically, but not in stratum 1.

Ignoring other phonological and morphological processes, the organization of the rules in question would be as shown in Figure (1). Some sample derivations (again ignoring other factors, in particular vowel quality) are shown in (2).

(1)	Phonological Rules	Morphological Rules
Lexicon:		
Level 1:	Nasal Assimilation (NA)	Comparative formation (CF)
Level 2:	Final -g deletion (FGD)	Compounding (CMP) Agentive formation (AF)

(2) (Note: vowel quality and other phonological processes ignored here)

UR:

	/long/	/long/ + /-er/	/anger/	/sing/ + /er/	/hang/ + /out/
Level 1:					
NA:	loNg	loNg	aNger	siNg	haNg
CF:	---	loNg+er	---	---	---
Output:	loNg	loNger	aNger	siNg	haNg
Level 2:					
FGD:	loN	---	---	siN	haN
CMP:	---	---	---	---	haN+out
AF:	---	---	---	siN+er	---
Output:	[loN]	[loNger]	[aNger]	[siNer]	[haNawt]

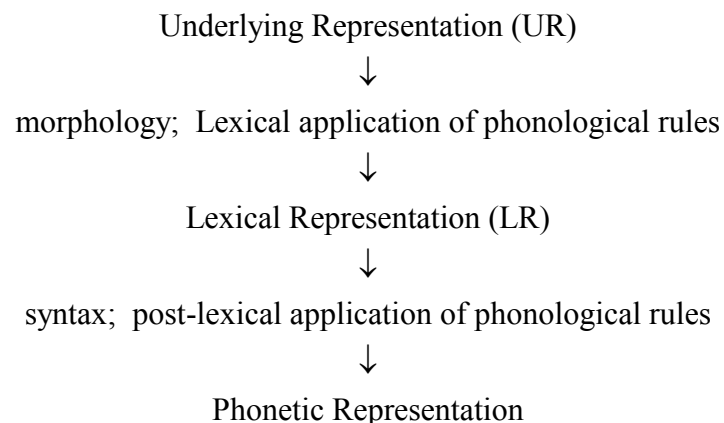
As the figure shows, the /g/ in morpheme-internal clusters (hunger, finger, anger, etc.) is never deleted because it is never final. The /g/ in clusters which precede the comparative -er

(longer, stronger, etc.) is final only entering stratum 1. By the time the rule of final /g/-deletion can apply (stratum 2), this /g/ is no longer final, and thus it is never deleted. But the /g/ in clusters which precede the agentive -er (singer, hanger, etc.), as well as in compound forms (long-eared, hangout, etc.), is final until stratum 2. If final /g/-deletion is ordered to apply before -er suffixation and compounding, the correct pattern is predicted.

The number of strata and the manner in which phonological rules are related to particular strata are points of contention between different versions of the theory (Kiparsky 1985; Mohanan 1986). But such differences will be irrelevant for our present purposes.

The organization of the LP model can be diagrammed as follows:

(3)



The term Lexical Representation (LR) refers to the output of the lexical component. In searching for analogies to earlier work, it is natural to think of UR as corresponding to morphophonemic representation and LR as corresponding to phonemic representation. Since morphophonemes were never as rigidly defined as phonemes, the former comparison seems unproblematic. However, there are important differences between LR and the classical phonemic level. As stated above, the most obvious difference relates to the biuniqueness condition of

descriptivist phonemics.

2. Classical phonemics and biuniqueness

In referring to the post-Bloomfieldians, there are two reasons to prefer the term “American Descriptivists” over the term used by Chomsky, “American Structuralists”. First, it seems helpful to distinguish this particular school from other structuralist approaches, both European and American, including Chomsky’s own. Secondly, it seems preferable (out of courtesy if nothing else) to use autonyms rather than exonyms wherever possible, and the term descriptivist seems to have been a term of self-reference for this group; witness the following infamous quote from Joos (1957):

An older term for the new trend in linguistics was ‘structural’. It is not idle to consider how the term ‘descriptive’ now came to replace it, even if not all the reasons can be identified. The Sapir way of doing things could be called structural, but the term was more often used for the stimulating new ideas that were coming out of Europe, specifically from the Cercle Linguistique de Prague. American linguistics owes a great debt to that stimulation; but in the long run those ideas were not found to add up to an adequate methodology. Trubetzkoy phonology tried to explain everything from articulatory acoustics and a minimum set of phonological laws taken as essentially valid for all languages alike, flatly contradicting the American (Boas) tradition that languages could differ from each other without limit and in unpredictable ways, and offering too much of a phonological explanation where a sober taxonomy would serve as well.

Children want explanations, and there is a child in each of us; descriptivism makes a virtue of not pampering that child. (p. 96)

The requirement of biuniqueness in phonemic representations was an innovation of the descriptivists. It was not characteristic of “the Sapir way of doing things”. Sapir’s guiding principle was that the phonemic representation should match the speaker’s intuitions (Sapir, 1933). Even Bloomfield did not require strict biuniqueness, which prompted the following comments from Joos (1957):

When we look at Bloomfield's work, we are disturbed at this and that, but more than anything else Bloomfield's confusion between phonemes and morphophonemes disturbs us. Bloomfield kept himself out of trouble here, usually, by describing just one language at a time, or one area within each at a time, adjusting for the effects of the confusion. But it made his procedure an unsafe model for neophytes and made the corpus of his work an inadequate source to distill procedural theory out of. (p. 92)

The final line is crucial to understanding the descriptivist approach. Their fundamental goal was to develop a procedural theory, i.e. a theory which would allow for operational definitions of linguistic concepts. This goal, while quite different from current attempts to derive linguistic principles by deductive reasoning from *a priori* assumptions, is not inherently unreasonable. However, the goal of "operational definition" of the phoneme turned out to conflict with the original concept of the phoneme as capturing speakers' judgements of identity vs. contrast.

The principle of biuniqueness was first stated by Bloch (1941). Bloch's goal was to make explicit the difference between partial overlap, i.e. two distinct phonemes sharing a common allophone in distinct environments, vs. complete overlap, in which two distinct phonemes share a common allophone in the same environment. Partial overlap was permissible in phonemic representations, but complete overlap could only be treated as morphophonemic variation between distinct phonemes.

The classic example of complete overlap concerns vowel length in Bloch's own dialect of American English. Bloch describes a productive allophonic rule which lengthens vowels before voiced consonants, as in:

(4)	beat [bit]	bead [bi:d]
	bit [bIt]	bid [bI:d]
	bet [bEt]	bed [bE:d]
	bat [bæt]	bad [bæ:d]

etc. The same rule applies to /a/, as in pot [pat] vs. pod [pa:d]. Thus Bloch states: “The vowel of pot is affected by the same automatic alternation...” However, in his dialect there are environments in which length is contrastive for /a/. Bloch cites a minimal pair, bomb [bam] vs. balm [ba:m]. Other examples of the short vowel are found in bother and sorry, in contrast to the long vowel in father, starry, pa, etc. This means that /a/ and /a:/ must be distinct phonemes. Bloch shows that this analysis would lead to assigning the same phonetic segment ([a:]) to different phonemes in the same environment: the [a:] in pod would be allophonically related to /a/, whereas the [a:] in pa'd would be a realization of /a:/. This violation of the biuniqueness principle forces Bloch to treat every occurrence of [a:] as belonging to the phoneme /a:/. If the length contrast between pot /pat/ and pod /pa:d/ is to be expressed as a phonological alternation, it can only be stated as a morphophonemic process, even though .

This points out a crucial difference between LR and the classical phonemic representation. The phonemic level was defined in terms of surface contrast -- any process that related two contrastive items had to be morphophonemic, whether or not the alternation was automatic. While the distinction between contrastive and non-contrastive features is important in LP, the distinction between automatic and non-automatic alternations is more fundamental in defining the level of Lexical Representation. Only lexical rules can have lexical exceptions. A rule which neutralizes a phonemic contrast (e.g. final devoicing in German) may never-the-less be post-lexical as long as it is automatic (i.e. exceptionless).

Bloch 's motivation for accepting this unsatisfying analysis was to preserve the transparency of “writing” rules, allowing the phonemic representation of an utterance to be absolutely predictable from phonetics alone. As he wrote in the last paragraph of his paper:

... by sacrificing this symmetry we are able to account for all the facts of pronunciation, which is surely the more important requirement. The resulting system is lopsided; but the classes it sets up are such that if we start from the actual utterance of the dialect we can never be in doubt of the class to which any particular fraction of the utterance must be

assigned.

3. The demise of the phoneme

Chomsky (1964) used Bloch's data as important evidence against the existence of a phonemic level. Of course, even to pose the question in these terms is to risk serious anachronism. The generative concept of a "level" as a stage in a derivation was totally foreign to the descriptivist approach. However, the point of Chomsky's criticism was this: data like that presented by Bloch calls into question the rationale for distinguishing between allophonic and morphophonemic rules.

Chomsky's underlying assumption is that the human mind seeks maximal generality. This implies that an optimal grammar is one which contains the least possible redundancy. Grammars should not contain two different rules which accomplish the same thing in the same environment (e.g. both an allophonic and a morphophonemic rule to lengthen vowels before voiced consonants). The fact that the biuniqueness requirement prevented Bloch from adopting the most economical (least redundant) analysis makes it unlikely (at least in Chomsky's view) that this requirement has any psychological reality.

Another key example cited by Chomsky was Halle's (1959) analysis of voicing assimilation in Russian. Halle points out that the process which changes [t] to [d], [p] to [b] and [k] to [g] before a voiced obstruent seems identical to the process which changes [c&] to [j&] in the same environment. However, voicing is contrastive for stops -- thus /p/ vs. /b/, /t/ vs. /d/ and /k/ vs. /g/ are all distinct phonemes. The voiced affricate [j&] on the other hand is not a phoneme in its own right; it arises only as an allophone of /c&/. Thus the rule changing /c&/ to [j&] is an allophonic rule, whereas the biuniqueness condition requires that the rule changing /t/ to /d/ etc.

be a morphophonemic rule.

Halle, like Chomsky, argued that the biuniqueness condition was not merely unnecessary but actually a hindrance to adequate phonological analysis, in that it forces the linguist to posit two different rules which produce the same changes in the same environment. Such arguments led generative phonologists to abandon not only the principle of biuniqueness but also the concept of phonemic representation. Schane (1971) pointed out that this was an undesirable development, arguing that phonemic distinctions could be determined on the basis of surface contrast. But he too denied the viability of a phonemic level as a unique stage in the derivation of surface forms.

As Schane and others pointed out, the output of the phonological rules in SPE-vintage generative work (Chomsky's "systematic phonetic level") was often virtually indistinguishable from a classical phonemic representation of the same data. But this was largely accidental rather than an inherent property of the theory. After Chomsky's attacks on complementary distribution as a criterion for phonological analysis, morphophonemic alternations became the primary data to be accounted for. Early generative work largely ignored the kind of phonetic detail typically observed in allophonic processes.

4. Lexical Representation vs. Phonemic Representation

The problem of Russian voicing assimilation provides a revealing comparison of the three theories. Descriptivist phonemics required two distinct rules, one allophonic and the other morphophonemic. The SPE framework required a single rule, denying the validity of the distinction between allophonic and morphophonemic rules. Lexical Phonology posits a single rule, but requires it to apply at two different times and in two different ways.

Within the lexical component, voicing assimilation can apply to only those segments for which voicing is contrastive (phonemic), because of the structure-preservation constraint. It applies cyclically and only within words. Post-lexically, however, voicing assimilation applies to all consonants and can apply across clitic or even full word boundaries. Moreover, while lexical voicing assimilation is categorial, Kiparsky cites evidence that the post-lexical application of the rule produces gradient effects, i.e. partial voicing (or devoicing) with the strongest effect observed in segments closest to the trigger. (The study of such post-lexical processes has been one of the factors contributing to a recent resurgence of interest in phonetic detail on the part of theoretical phonologists.)

Thus the LP analysis agrees with the classical phonemic analysis in claiming that the alternation between /t/ and /d/ has a different status than the alternation between /c&/ and [j&]. Kiparsky (1985, p. 113-114) notes:

Since Russian Voicing Assimilation has figured so prominently in the debate on the phonemic level, a final remark on the relationship between our lexical representations and the structuralist phonemic representations may be appropriate. With regard to the classic point of contention itself, our position should satisfy both parties. The output of the lexical phonology contains of course the voiced obstruents /b d g z &.../, but it does not contain the voiced allophones of the phonemes that lack a phonemic voiced counterpart, namely /c&/, /c/, /x/, But neither do we require two separate Voicing Assimilation rules: we have a *single* rule which applies both lexically and postlexically with different results as dictated by the principles of the theory.

5. On “grammatical prerequisites”

Another difference between the descriptivist phonemic representation and Lexical Representation is the descriptivists’ insistence on the strict separation of phonological and grammatical analysis. Specifically, the descriptivists insisted that phonemic analysis could make no reference to grammatical structure. This view, on the face of it, seems to be the antithesis of the position adopted in LP; yet even here there are similarities. Recall that LP requires the post-

lexical phonology (corresponding to the allophonic rules which a phonemic analysis would identify) to be blind to word-internal morphological structure.

Pike (1947, 1952) and Jakobson (1948) argued against the doctrine of separatism, and Jakobson cited similar arguments by Sapir from 25 years earlier. However, the goal of operationalizing the definition of linguistic units drove the descriptivists to reject these arguments, for fear that mixing phonological and grammatical information would lead to circular definitions.

Again, it is important to distinguish between the analytical procedures of the linguist and the formal presentation of the results of that analysis. The descriptivists recognized that every linguist deals with grammatical and phonological facts simultaneously, using each to shed light on the other. But they felt that a rigorous description of the linguistic facts could only be non-circular by proceeding from the bottom up, i.e. from smallest to largest units, with description on each level making no reference to units on a higher level (cf. Hockett 1942, Bloch 1950).

Pike's papers came under attack, both directly (as in subsequent papers by Hockett, Bloch, Trager and Wells) and indirectly, as in Joos's comments on the paper by Bloch (1941) discussed above:

It was the present article by Bloch that made clear, as it never had been before, that phonemics must be kept unmixed from all that lies on the opposite side of it from phonetics ... It was a great deal to accomplish in so few pages, and few readers realized that it had been done: the ghost of the slain dragon continued to plague the community of linguists under such names as 'grammatical prerequisites to phonemic analysis' and has not been completely exorcized to this day. (Joos 1957, p. 96)

Pike's discussion of the "interweaving of grammatical and phonemic facts" is at times strongly reminiscent of the LP conception of the interleaving of morphology and phonology. But of course Pike did not distinguish between lexical and phrasal processes. He was equally

concerned with phrasal phenomena, e.g. the role of grammatical boundaries as potential boundaries for intonational contours, as with word-level processes.

Pike raised several other points that foreshadowed later theoretical developments in interesting ways. He argued in favor of using alternations in morpheme shapes as “clues” to allophonic relationships, and against positing quasi-segmental “juncture” phonemes -- an issue of some interest in post-SPE generative phonology. Pike argued instead that the boundaries of grammatical constituents should be recognized as potentially conditioning phonological processes, which is quite close to the position adopted by LP.

6. Conclusion

Lexical Phonology represents an integrated approach to the analysis of morphological and phonological structure which recaptures a number of the important intuitions of classical phonemics. The output of the lexical component corresponds closely to classical phonemic representation. The main difference between the two (namely the non-biuniqueness of the former) results from the fact that post-lexical rules in LP can neutralize phonemic contrasts, so long as these rules are exceptionless. But the notion of “phonemic contrast” does play a significant role in LP, e.g. in the formulation of the Structure Preservation Principle as prohibiting lexical operations on non-contrastive features. In fact, the word “phoneme” itself is no longer taboo in generative phonology.

References

- Bloch, Bernard. 1941. Phonemic overlapping. *American Speech* 16.278-84.
- Bloch, Bernard. 1950. Studies in colloquial Japanese IV: Phonemics. *Language* 26.86-125.

- Chomsky, Noam. 1964. Current issues in linguistic theory. Mouton: The Hague.
- Halle, Morris. 1959. The Sound Pattern of Russian. Mouton: The Hague.
- Hockett, Charles F. 1942. A system of descriptive phonology. *Language* 18.3-21.
- Jakobson, Roman. 1948. The phonemic and grammatical aspects of language in their interrelation. Proceedings of the Sixth International Congress of Linguists, 5-18.
- Joos, Martin, ed. 1957. Readings in Linguistics I: The development of descriptive linguistics in America 1925-56. (4th edition, 1966) Chicago: University of Chicago Press.
- Kaisse, Ellen M. and Patricia A. Shaw. 1985. On the theory of Lexical Phonology. Phonology Yearbook 2, 1-30.
- Kiparsky, Paul. 1982. Lexical phonology and morphology. In I.-S. Yang (ed.), Linguistics in the Morning Calm. Seoul: Hanshin. 3-91.
- Kiparsky, Paul. 1985. Some consequences of lexical phonology. Phonology Yearbook 2, 83-138.
- Mohanan, K.P. 1986. The Theory of Lexical Phonology. Dordrecht: Reidel.
- Mohanan, K. P. and Tara Mohanan. 1984. Lexical Phonology of the consonant system in Malayalam. *Linguistic Inquiry* 15:575-602.
- Pike, Kenneth L. 1947. Grammatical prerequisites to phonemic analysis. *Word* 3.155-72.
- Pike, Kenneth L. 1952. More on grammatical prerequisites. *Word* 8.106-121.
- Sapir, Edward. 1925. Sound patterns in language. *Language* 1.37-51.
- Sapir, Edward. 1933. The psychological reality of phonemes. In Mandelbaum, David G. (ed.), Selected Writings of Edward Sapir in Language, Culture and Personality. Berkeley and Los Angeles: Univ. of California Press.
- Schane, Sanford A. 1971. The phoneme revisited. *Language* 47.503-521.